 NOT TO BE FURTHER COPIED OR REPRODUCED

(* y VS x *)

13

Too many points?

(cutting down a larger example)

Introduction 13-1

13A. A floral example 13-1

13B. The other line 13-3

13C. The joint schematic plot 13-7

13D. Thinning 13-16

13E. A thin trial tilt 13-27

13F. Interlaced thinning (optional techniques) 13-29
- computer version, 13-31

13G. Looking at residuals 13-35
- working with x-arrays, 13-40

* * *

13H. How far have we come? 13-44

* * *

13K. Using spread-out values (optional technique) 13-45

13L. Still more points? (optional background) 13-46
- more variables, 13-46
- even more points, 13-51

* * *

13P. Problems 13-54

Exh.		Page
1	A	13-2
2		13-4
3		13-5
4		13-6
5	B	13-8
6		13-9
7	C	13-10/1[illegible]
8		13-13
9		13-14
10		13-15
11		13-17
12	D	13-19/2[illegible]
13		13-22
14		13-23/2[illegible]
15		13-26
16		13-28
17	E	13-30
18	F	13-32/3[illegible]
19		13-34
20		13-36
21	G	13-37
22		13-38
23		13-39
24		13-41
25		13-43
26	H K	13-47/-[illegible]
27		13-49/5[illegible]
	L P	

13

Too many points?

(cutting down a larger example)

We have so far treated not very large batches of (x,y) points. No doubt these seemed to take enough effort.

As we go to medium or medium large batches, we need to introduce new techniques if we are going to find either hand analysis or useful plotting possible. This chapter illustrates techniques that can help.

13A. A floral example

Exhibit 1 gives information about number of pistils and number of stamens for 268 early flowers of *Ranunculus ficaria*. For both compressed presentation and easy manipulation, the data are given in the form of a two-way table of counts. Here the entry "7" under 27 stamens and opposite 18 pistils means that 7 of the 268 flowers had exactly 27 stamens and 18 pistils.

If we had data for fewer flowers, we would proceed with the individual values, dividing into three portions in the usual way. We have enough observations -- more precisely there is enough repetition of the individual values of x -- to gain by first summarizing what happens for fixed x and then working with these summaries.

Numbers of stamens and pistils for 268 early flowers of

Rununculus ficaria

A) THE DATA

Number of pistils, y	18	20	22	24	26	28	30	32	34	36	37	38	(Sum)
	Number of stamens, x												
≤6											(37,2)		(1)
=7							1						(1)
8													
							1						(1)
10		1			1								(2)
			2		1								(3)
12		3	1	2	1	1	3	1	1				(13)
		1	1	1	4	1	1	1		1	1		(12)
14		4	3		1	2	4	1	2	3	1	1	(22)
			1	2	4	3	7	4	4	5	2	1	(35)
16	1			2			1	5	3	5	4	5	3	2			(31)
					2	2	1	4	2	3	1	2	5	2	1		(25)
18				1		2	4	3	1	7	1	3	2	1		2	(27)
						1	2		5	4	4	1	1	1	1	1	(21)
20					2		1	1	2	2		1	3	4	2	1	(19)
					1				2	2			2	4	1	1	(13)
22								1	2	3		1	2	3	1	2	(15)
									1	1		2	1	1	2		2	(10)
24									1	1		2						(4)
														1	2			1	(4)
26													1		2				(3)
								1		1							2		(4)
28																1			(1)
30																			
											1								(1)
32																			
(Sum)	1,	6,	8,	9,	16,	12,	22,	26,	26,	38,	14,	23,	20,	20,	13,	7,	1,	4,	-,	1,	1	268

B) SOURCE

L. H. C. Tippett 1952. The methods of statistics 4th edition. New York, John Wiley.

Exhibit 2 shows the detailed calculations. We have taken medians for x-arrays with 4 or fewer observations. As a first calculation, we will treat the 15 trimeans as if they were individual observations of equal importance. Exhibit 3 shows the calculation of an apparent tilt of 0.51 and the application of a trial tilt of 0.50. Clearly

$$0.5x + 4$$

is a good first approximation to the fit.

With this in hand, we can reconsider our treatment of all the trimeans as equally important. The natural thing to do is to give each trimean (or median) a weight equal to the number of observations that entered into it. (We could have done this from the beginning, but would have gained nothing from the extra effort.) Exhibit 4 shows the results of doing this and fitting a further tilt, which turns out to be +0.05. The final fit is, with a further adjustment to level of +0.1-1.3=-1.2

number of pistils = 0.55(number of stamens) + 2.8

Clearly the single observations with x = 18, 34,35, and 37 are suspiciously discrepant.

13B. <u>The other line</u>

If we interchange x and y, we obtain the results shown in exhibits 5 and 6. Making a further adjustment of 0.02y gives the following summary residuals

old	-0.2	0.2	0.0
0.02y	.26	0.34	0.44
diff	-0.46	-0.32	-0.44

exhibit 2 of chapter 13

Summarization of x-arrays

for y

		Depth for		Values of			
x	**Count**	**H**	**M**	**H**	**M**	**H**	**tri**
18	1		1		16		
19	6	2	3.5	13	14	14	13.8
20	8	2.5	4.5	12	13.5	14	13.2
21	9	3	5	12	15	16	14.5
22	16	4.5	8.5	13	15	17	15.0
23	12	3.5	6.5	14	15	17.5	15.4
24	22	6	11.5	14	15	18	15.5
25	26	7	13.5	14	16	17	15.8
26	26	7	13.5	15	18.5	20	18.0
27	38	10	19.5	16	18	20	18.0
28	14	4	7.5	15	16	19	16.5
29	23	6	12	16	17	22	18.0
30	20	5.5	10.5	17	18.5	21	18.8
31	20	5.5	10.5	17.5	20	21.5	19.8
32	13	4	7	20	22	25	22.2
33	7	2	4	18	20	22	20.0
34	1		1		28		
35	4		2.5		25		
36	.		.		.		
37	1		1		2		
38	1		1		25		

exhibit 3 of chapter 13

Fitting of tilt to the 15 trimeans
(center values are trimeans of 5)

x	y-trimean	center values for x	center values for y	0.5x	Diff*	Center
19	13.8			9.5	4.3	
20	13.2			10	3.2	
21	14.5	21	14.4h	10.5	4.0	4.0
22	15.0			11	4.0	
23	15.4			11.5	3.9	
24	15.5			12	3.5	
25	15.8			12.5	3.3	
26	18.0	26	16.7	13	5.0	3.7
27	18.0			13.5	4.5	
28	16.5			14	2.5	
29	18.0			14.5	3.5	
30	18.8			15	3.8	3.8h
31	19.8	31	19.6	15.5	4.3	
32	22.2			16	6.2	
33	20.0			16.5	3.5	

Apparent tilt = $\frac{19.6 - 14.45}{31-21}$ = 0.515

Trial tilt = 0.5

Diff* = (y-trimean) - 0.5x.

exhibit 4 of chapter 13

Weighted analysis of trimeans (or medians)

(see exhibits 2 and 3)

x	wt	(for x)	y-0.5x-4			(*)	Reordered by residual within thirds	
18	1		3.0	-0.8	to	8		
19	6	74	0.3	-0.5		30	74	
20	8	19H21	-0.8	-0.1	to	42	19H	-0.5
		37hM22					37hM	-0.1
21	9	19H24	0.0	0.0	to	32 ✓	19H	0.0
22	16	tri22.2	0.0	0.0		23	tri	-0.2
23	12		-0.1	0.3		7		
24	22		-0.5	3.0	to	1		
25	26		-0.7	-1.5	to	14		
26	26	127	-1.0	-0.7		40	127	
27	38	32H26	0.5	-0.5	to	63 ✓	32H	-0.7
		64M27					64M	0.5
28	14	32H28	-1.5	0.5	to	64	32H	0.5
29	23	tri27	-0.5	1.0	to	26	tri	0.2
30	20		-0.2	-20.5	to	1		
31	20		0.3	-0.5		8		
32	13	67	2.2	-0.2	to	28 ✓	67	
33	7	17H30	-0.5	0.3	to	39	17H	-0.2
34	1	34M31	7.0	2.0		9	34M	0.3
35	4	17H32	3.5	2.2		8	17H	0.3
36	·	tri31		3.5		5	tri	0.2
37	1		-20.5	7.0	to	1		
38	1		2.0					

Apparent tilt = $\frac{(0.2) - (-0.2)}{31 - 22.2} = 0.05$

Apparent level = $\frac{(-0.2) + (0.2) + (0.2)}{3} = 0.1$

Adjustment to level for apparent tilt = $-\frac{1.11 + 1.35 + 1.55}{3} = -1.34$

*Examples: Wt of -0.8 = 8, Wt of -0.5 = 22, 8+22 = 30, Wt of 0.1 = 12, 30+12 = 42, etc.

and since

$$\frac{1}{3}\Big((-0.46) + (-0.32) + (-0.44)\Big) = -.41, \quad \text{say } -.4$$

the final fit is

number of stamens = 0.77(number of pistils) + 13.1

Clearly the single observations with y = 2,7,9 and 31 are suspiciously discrepant.

13C. <u>The joint schematic plot</u>

Exhibit 7 calculates the eighths and hinges for x, y, x + y, and x - y -- with the basic data for the last two obtained from exhibits 8 and 9 -- and then goes on to identify outside and detached points.

Exhibit 10 shows:

-middle oval.

-detached points.

-outside points.

-outer polygon (using the outermost points neither detached or outside, points which can be easily identified from exhibit 1).

We see a rather blunt oval, whose shape is reinforced by the outer polygon and points. The detachment of 1 of the 3 detached points is very obvious. The other 2 are far from the body of the batch, but may or may not be of a distinctively different character.

exhibit 5 of chapter 13

First steps with y and x interchanged

(central values are medians)

y	x-trimean			0.75y	Diff	
12	23.2			9	14.2	
13	22.5	13.5	23.4	9.8	12.7	13.4
14	23.5			10.5	13.0	
15	25.0			11.2	13.8	
16	27.0			12	15.0	
17	27.2	17.5	27.0	12.8	14.4	14.0
18	27.0			13.5	13.5	
19	27.0			14.2	12.8	
20	29.2			15	14.2	
21	29.5	21.5	29.5	15.8	13.7	13.5
22	29.5			16.5	13.0	
23	30.5			17.2	13.3	

$$\text{Apparent tilt} = \frac{29.5 - 23.4}{21.5 - 13.5} = .76$$

Trial tilt = 0.75

exhibit 6 of chapter 13

Weighted analysis with x and y interchanged

y	wt	cum	for y		x-.75y-13.5	Resid	wt	cum		
2	1	1			22.0	-0.8	12	12		
7	1	2			11.3	-0.7	3 to	15		
9	1	3	55		9.8	-0.5	222 ✓	22	55	
10	2	5	14H	12	1.0	0.7	13 to	18	14H	-0.7
11	3	8	28M	13	-0.7	1.0	2	5	28M	-0.5
12	13 to	21	14H	14	0.7	9.8	1	3	14H	0.7
13	12 ✓	12	tri	13	-0.8	11.3	1	2	tri	-0.2
14	22 to	22			-0.5	22.0	1	1		
15	35	35			0.1	-0.7	21 to	21		
16	31 to	66	139		1.5	-0.0	27 to	48	139	
17	25 ✓	25	35H	15	0.9	0.1	35 ✓	35	35H	0.0
18	27 to	48	70M	17	0.0	0.9	25 to	56	70M	0.1
19	21	21	35H	18	-0.7	1.5	31	31	35H	0.9
			tri	17.2					tri	0.2
20	19	19			0.7	-5.7	1 to	1		
21	13 to	32			0.2	-3.5	4	5		
22	15	15			-0.5	-2.9	4	9		
23	10 to	27	74		-0.2	-0.6	1	10	74	
24	4	17	19H	20	-3.5	-0.5	15	25	19H	-0.5
25	4	13	37hM	22	0.7	-0.2	10 to	35	37hM	0.0
26	3	9	19H	23	0.0	0.0	✓	3	19H	0.7
27	4	6	tri	21.8	-2.9	0.2	13 to	36	tri	0.0
28	1	2			-0.5	0.7	19	23		
31	1	1			-5.7	0.7	4	4		

Apparent 2nd tilt = $\frac{0.0 - (-0.2)}{21.8 - 13.0}$ = .02

exhibit 7 of chapter 13

Calculation of, and with, eighths and hinges
for x, y, x+y, x-y

(Count = 268, E has depth 34, H has depth 67h)

A) CALCULATION FOR x

x	Low depths	#	x	High depths	#	Out/det	
18	1		38	1	0		
19	2 to 7		37	2	0	T	13h
20	8 to 15		35	3 to 6		S	16h
21	16 to 24		34	7		34	22
22	25 to 40 E		33	8 to 14		67h	24
23	42 to 52		32	15 to 27		67h	29h
24	53 to 74 H		31	28 to 47 E		34	31
25			30	48 to 67		S	36h
			29	68 to 90 H		T	40
						dh	5h
						dl	9

B) CALCULATION FOR y

y	Low depths	*	y	Low depths	#	Out/det	
2	1	D	31	1	D		
7	2	0	28	2	0	T	6
9	3		27	3 to 6		S	9
10	4 to 5		26	7 to 9		34	14
11	6 to 8		25	10 to 13		67h	15
12	9 to 21		24	14 to 17		67h	20
13	22 to 33		23	18 to 27		34	
14	34 to 55 E		22	28 to 42 E		S	27
15	56 to 90 H		21	43 to 55		T	30
16			20	56 to 74 H		dh	5
17			19			dl	8

exhibit 7 (Cont'd)

C) CALCULATION FOR x+y

x+y	Low depths	(*)
29	1	0
32	2 to 7	
33	8 to 13	
34	14 to 20	
35	21 to 27	
36	28 to 34 E	
37	35 to 46	
38	47 to 56	
39	57 to 69 H	

x+y	High depths	(*)	Out/det
63	1	0	
62	2 to 4	0	
60	5		T 20 S 27
58	6 to 9		34 36
57	10 to 11		67h 39 67h 48
56	12 to 13		34 52
55	14 to 17		S 61 T 68
54	18 to 21		dh 9 dl 16
53	22 to 29		
52	30 to 41 E		
51	42 to 52		
50	53 to 58		
49	59 to 66		
48	67 to 75 H		

D) CALCULATION FOR x-y

x-y	Ranks	(*)
-2	1 to 2	0
0	3	
1	4	
2	5 to 8	
3	9 to 12	
4	13 to 18	
5	19 to 39 E	
6	40 to 61	
7	62 to 81 H	

x-y	Ranks		
35	1	D	
22	2	D	
20	3	O	T -3
17	4		S 0
16	5 to 6		34 5 67h 7
15	7 to 15		67h 12 34 13
14	16 to 25		S 18
13	26 to 47	E	T 21 dh 5
12	48 to 70	H	dl 8
11	71 to 94		

exhibit 7 (cont'd.)

E) Points OUTSIDE or DETACHED

By ordinale

	(outside only)	(detached)
Low x	----	---
High x	(37,2),(38,25)	---
Low y	(29,7)	(37,2)
High y	(34,28)	(29,31)
Low x+y	(10,19)	---
High x+y	(34,28),(35,27),(35,27), (38,25)	---
Low x-y	(29,31),(25,27)	---
High x-y	(29,9)	(29,7),(37,2)

Consolidated list

Detached: (29,7),(29,31),(37,2)

Outside only: (10,19),(25,27),(29,9),(34,28),(35,27), (35,27),(38,25)

exhibit 8 of chapter 13

counting-up occurrences of x+y

x

x+y	18	20	22	24	26	28	30	32	34	35	37	38	(Sum)
29		1											(1)
30													-
31													-
32		1 3	2										(6)
33		4 1	1										(6)
34	1	3	1 2										(7)
35		1	4	1	1								(7)
36			2 1	1 1	1		1						(7)
37			2 4	2 1	3								(12)
38				3 4	1 1		1						(10)
39			1 2	7	1	1					1		(13)
40				2 1	4 2								(9)
41				2 1	5 4	3 1							(16)
42			2	1 4	4 3	5 1	1						(21)
43			1	2	3 2	5 2	1						(16)
44				1	1	3 4	1						(10)
45					1 5	7 1	5						(19)
46					2	4 1	2 3	1					(13)
47					1 2	2 4	3 5	2 1					(20)
48					2	2	1 2	2					(9)
49					1	3	1 1	1 1					(8)
50					1	1	3	1					(6)
51						1	1 2	4 1	2				(11)
52					1		2 2	4 2	1				(12)
53							2 1	3 1	1				(8)
54						1		1 1	1				(4)
55								2	2				(4)
56							1	1					(2)
57								2					(2)
58								2		2			(4)
59													-
60							1						(1)
61													-
62									1	2			(3)
63												1	(1)

268, check

exhibit 9 of chapter 13

counting-up occurrences of x-y

x-y	18	20	22	24	26	28	30	32	34	35	37	38	(Sum
low													
-2					1		1						(2)
-1													
0						1							(1)
1			1										(1)
2	1		2		1								(4)
3			1		1	1							(4)
4				1 1	2	1	1						(6)
5	4	1	2 2	2 2	1 2	3	2						(21)
6	1	3	2	2 4	2	2	2	1 2	1				(22)
7		1	4	1	3 5	2	1	1 2					(20)
8		3	1 1	3 1	4 1	4	2	1		2			(23)
9	1		1 4	2 7	5 2	7 4	1 2	3 2					(41)
10			2 2	1 4	4 3	3 1	1 3	4 1					(29)
11				1 1	1 4	5 1	3 1	4 1	2				(24)
12				1	1 2	5 4	2 2	1 2	1	2			(23)
13					3	3 2	5 5	1 1	1			1	(21)
14					1 1	1	1 3	2	1				(10)
15					1	1 1	1	2	1	2			(9)
16							1	1					(2)
17								1					(1)
18													
high							2				1		(3)

		(x-y)	260. check
High:	(29,7)	22	
	(29,9)	20	
	(37,2)	35	

exhibit 10 of chapter 13

Joint Schematic Plot for the Floral Example

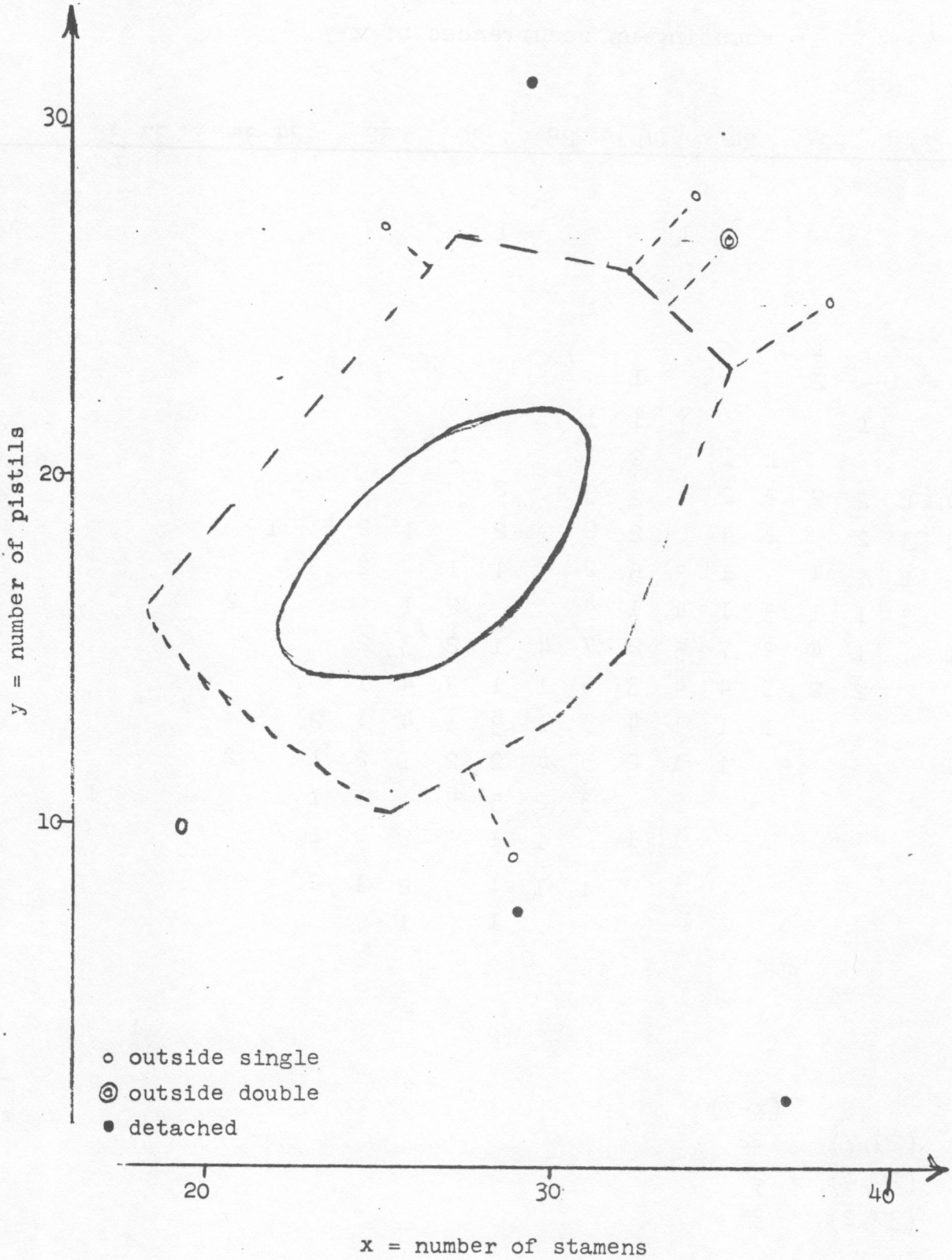


Exhibit 11 shows the two fitted lines (sections 13A and 13B) together with the schematic oval and polygon.

13D. Thinning

We don't want to have to confine our looking at batches of points to looking at schematic plots. We will often do very well with the schematic plots alone, but we should be able to look at the points themselves when this can help. Yet if we have many points, plotting them may require much effort and produce unsatisfactory results. How to escape?

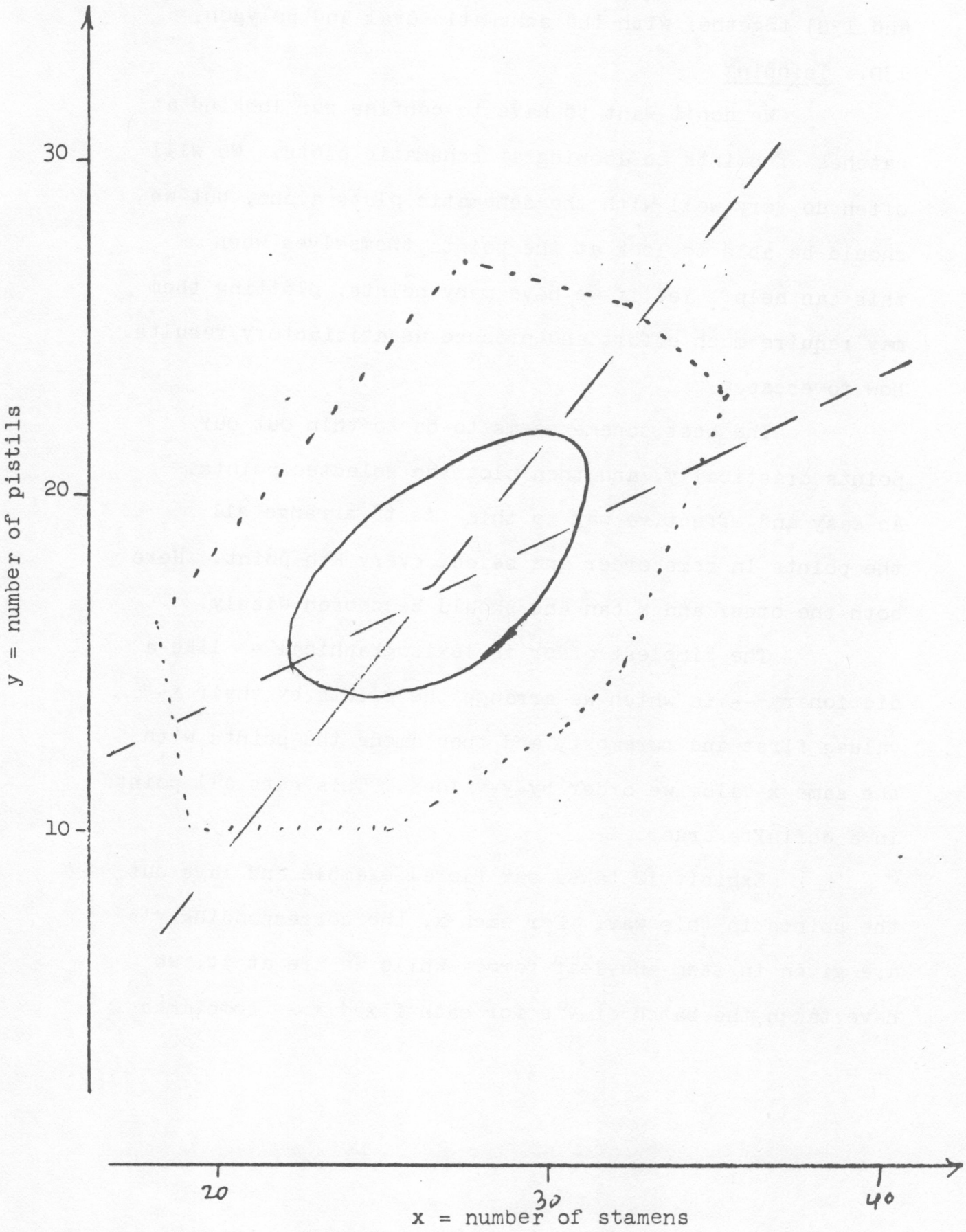
The best scheme seems to be to thin out our points drastically, and then plot the selected points. An easy and effective way to thin is to arrange all the points in some order and select every kth point. Here both the order and k can and should be chosen wisely.

The simplest order is lexicographical -- like a dictionary -- in which we arrange the points by their x-values first and foremost, and then among the points with the same x-value we order by y-values. This gets all points in a definite order.

Exhibit 12 takes our floral example and lays out the points in this way. For each x, the corresponding y's are given in stem-and-leaf form. While we are at it, we have taken the batch of y's for each fixed x -- sometimes

exhibit 11 of chapter 13

Fitted lines and schematic plot for the floral example



called an x-array -- and repeated the summarization by trimeans of exhibit 2, thus illustrating how one layout can serve both thinning and line fitting.

In panel A, every 10th point -- beginning with the 4th -- is indicated by a dark leaf. The 27 points thus obtained are listed in panel B.

Exhibit 13 shows these 27 points against a background of the middle oval and outer polygon of exhibit 11, obtained above from all 268 points. We do get some feeling of compatibility for the pattern of 27 points, and the schematic plot based on all 268.

The most natural thing to do as a first check on matters is to interchange the roles of y and x, sorting principally on y, and sorting on x only within y = constant. Exhibit 14 mirrors exhibit 12 in just this way, starting with the 6th point, not the 4th.

Exhibit 15 presents the new set of 27 points against the same background as exhibit 51. Again we get general concurrence.

We have not received much new information from adding 27 points, but we have received considerable reinforcement. After all, should we have expected more from only 27 points?

exhibit 12 of chapter 13

Data and summarization by number of stamens (=x) of the number of pistils (=y) in each of 268 early flowers of Ranunculus Ficaria

(see exhibit 1)

A The DATA -- every 10th leaf bold

x	y	#	H	M	H	trimean*	(outside y's)
18	1|6	1		16		16	
19	1|034444	6	13	14	14	13.8	(10)
20	1|2223444 1|5	8	12	13.5	14	13.2	
21	1|1123 1|55668	9	12	15	16	14.5	
22	1|2233334 1|555577 2|001	16	13	15	17	15.0	
23	1|2344 1|55577889	12	14	15	17.5	15.4	
24	1|234444 1|5555555607888899 2|0	22	14	15	18	15.5	
25	1|0122234 1|55556666677777888 2|027	26	14	16	17	15.8	(10,22,27)
26	1|244 1|5555666778999999 2|00112234	26	15	18.5	20	18.0	

exhibit 12 (Cont'd)

x	y	#	H	M	H	trimean*	
27	1|2444 1|5555566666777888888999 9 2|001122234 2|7	38	16	18	20	18.0	(27)
28	1|34 1|556666789999	14	15	16	19	10.5	
29	0|79 1|34 1|566666778889 2|023344 2| 3|1	23	16	17	22	18.0	(7,9)
30	1|66677777889 2|00011223 2|6	20	17	18.5	21	18.8	(26)
31	1|5667789 2|000011112223 2|5	20	17.5	20	21.5	19.8	
32	1|579 2|001233 2|5566	13	20	22	25	22.2	
33	1|889 2|0122	7	18	20	22	20.0	
34	2|8	1	--	28	--	28.0	
35	2|3377	4	23	25	27	25.0	
36	-----		----------				
37	0|2	1	--	2	--	2.0	
38	2|5	1	--	25	--	25.0	

*Median for batches of 3 or 4, of course.

exhibit 12 (Cont'd)

B The SELECTED POINTS -- in (x,y) form

(19,14), (20 14), (21,18), (22,15), (23,14), (24,13), (24,15), (24,20), (25,15), (25,17), (26,15), (26,19), (26,22), (27,15), (27,18), (27,19), (27,27), (28,18), (29,16), (29,19), (30,16), (30,20), (31,16), (31,21), (32,19), (32,26), (35,23).

13-22

exhibit 13 of chapter 13

combined approach with thinning

27 selected points and joint schematic plot

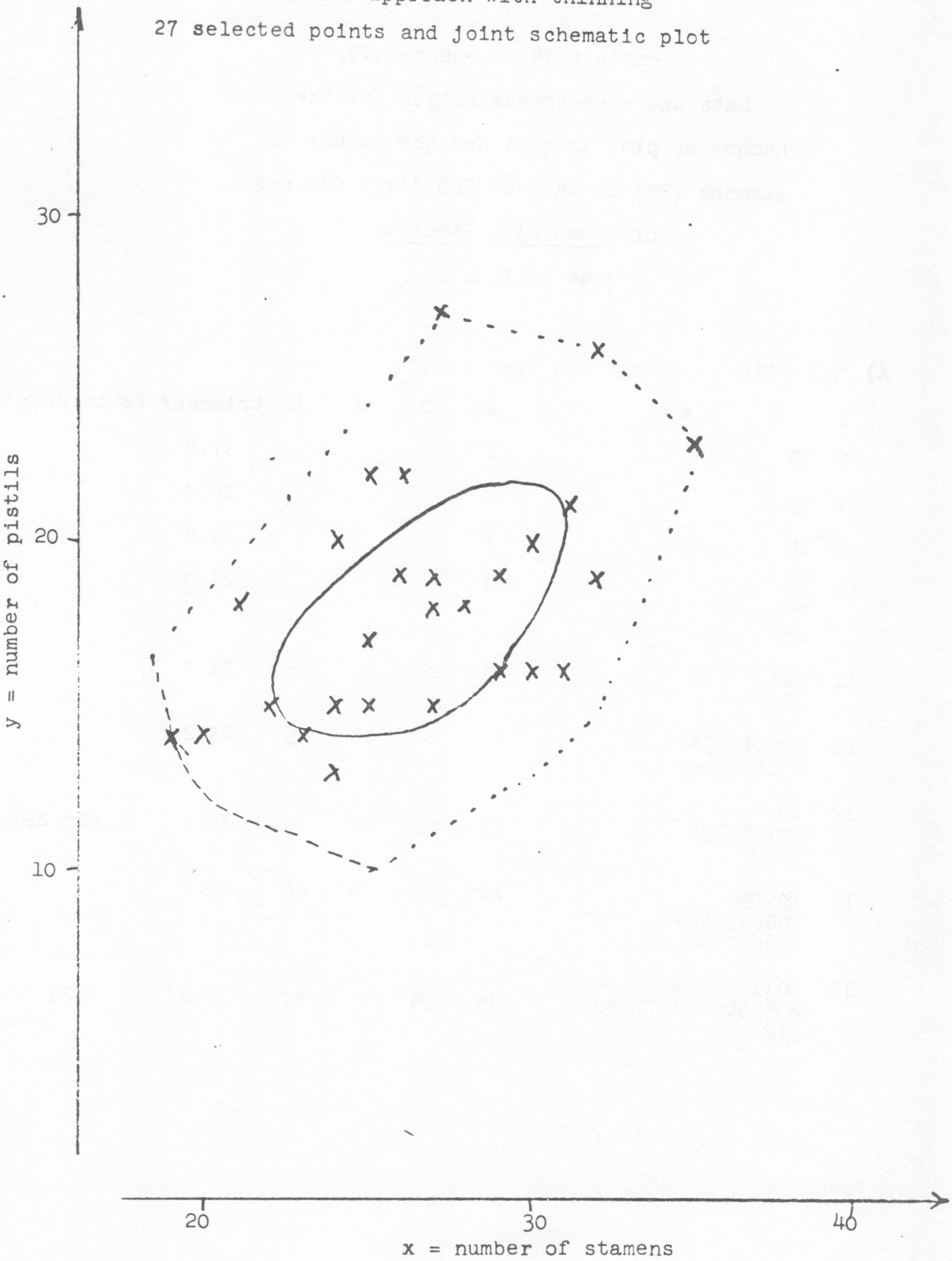


exhibit 14 of chapter 13

Data and some summarization for the number of pistils (=y) and the number of stamens (=x) in each of 268 early flowers of <u>Ranunculus Ficaria</u>

(see exhibit 1)

A) THE DATA -- every 10th leaf bold

y	x	#	H	M	H	trimean*	(outside y's
2	3|7	1				37.0	
7	2|9	1				29.0	
9	2|9	1				29.0	
10	1|9 2| 2|5	2				22.0	
11	2|11 2|5	3	21	21	25	21.0	
12	2|00012234 2|55567	13	22	23	25	23.2	
13	1|9 2|01222234 2|589	12	21h	22	24h	22.5	(28,29)
14	1|9999 2|0002334444 2|56677789	22	20	24	26	23.5	
15	2|01122223334444444 2|5555666677777889 3|12	35	23	25	27	25.0	(32)

exhibit 14 (Cont'd)

y	x	#	H	M	H	trimean*	(outside y's)
16	1|8 2|114 2|555556667777788889999 9 3|00011	31	25	27	29	27.0	(18)
17	2|22334 2|555566777899 3|00000112	25	25	27	30	27.2	
18	2|1334444 2|555677777778999 3|00133	27	25	27	29	27.0	(33,33)
19	2|344 2|66666777788889 3|0123	21	26	27	28	27.0	(23,31,32,33)
20	2|224 2|566779 3|0001111223	19	26	30	31	29.2	
21	2|2 2|6677 3|00111123	13	27	30	31	29.5	(22)
22	2|5667779 3|00111233	15	27	30	31	29.5	
23	2|6799 3|0122 3|55	10	29	30h	32	30.5	
24	2|6799	4	26.5	28	29	28.0	
25	3|122 3|8	4	31.5	32	35	32.0	
26	3|022	3	30	32	32	32.0	
27	2|57 3| 3|55	4	26	31	35	31.0	

exhibit 14 (Cont'd)

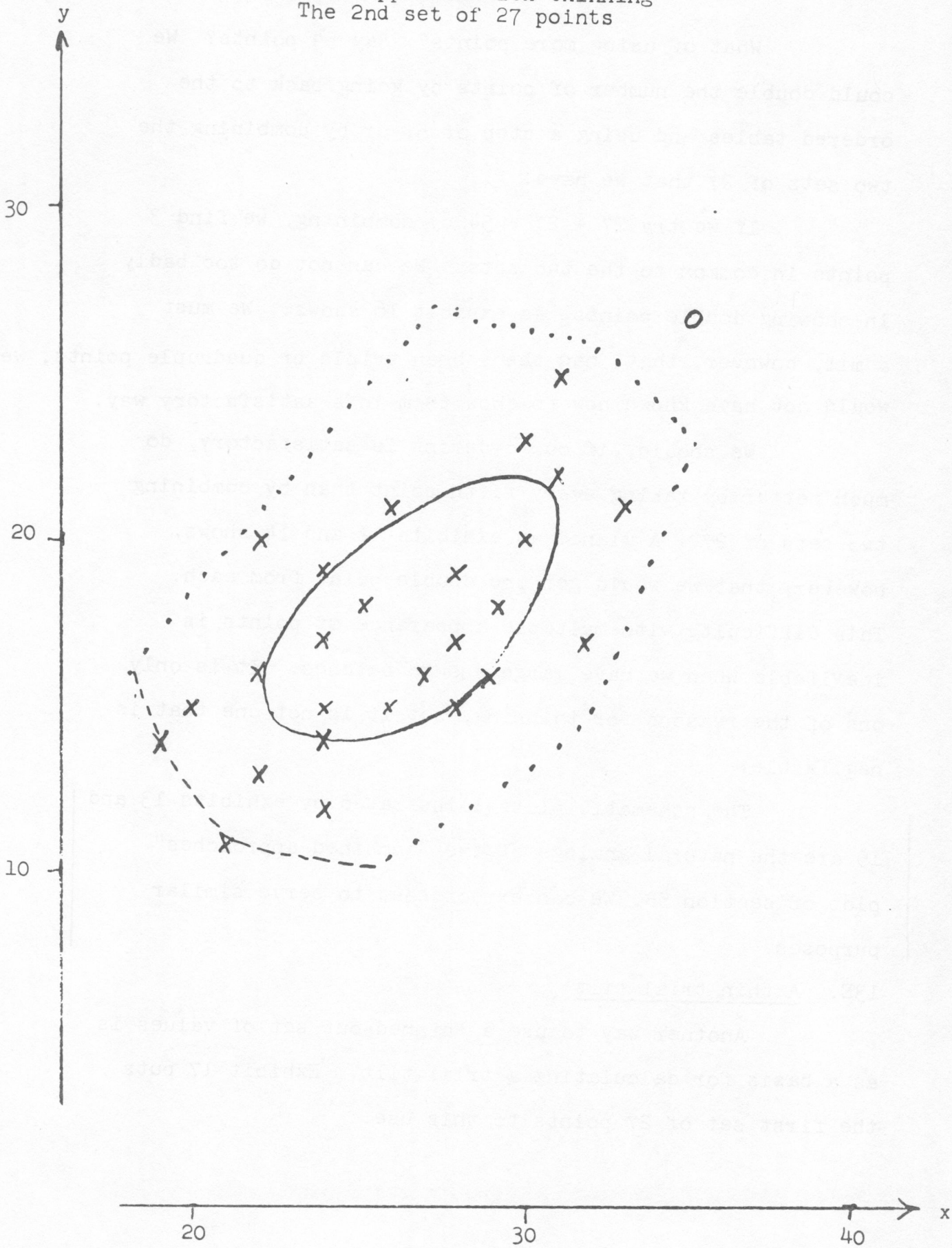
y	x	#	H	M	H	trimean*
28	34	1				34.0
31	29	1				29.0

*Median for batches of 3 or 4, of course.

B) THE SELECTED POINTS -- in (y,x) form

(11,21), (12,24), (13,22), (14,19), (14 24), (15,20), (15,24), (15,25), (15,28), (16,25), (16,27), (16,29), (17,24), (17,28), (17,32), (18,25), (18,29), (19,24), (19,28), (20,22), (20,30), (21,26), (21,33), (22,31), (23,30), (25,31), (27,35).

exhibit 15 of chapter 13
Combined approach with thinning
The 2nd set of 27 points



What of using more points? Say 54 points? We could double the number of points by going back to the ordered tables and using a step of 5, or by combining the two sets of 27 that we have.

If we try 27 + 27 = 54 by combining, we find 3 points in common to the two sets. We can not do too badly in showing double points, as exhibit 16 shows. We must admit, however, that, had there been triple or quadruple points, we would not have known how to show them in a satisfactory way.

We should, if our ordering is satisfactory, do much better by taking every fifth point than by combining two sets of 27. A glance at exhibits 12 and 14 shows, however, that we would get one double point from each. This difficulty with multiple appearance of points is inevitable when we have large enough batches. It is only one of the reasons for thinning, but it is not one that is negligible.

The schematic plots illustrated by exhibits 13 and 15 are the natural analogs of the "combined-approaches" plot of section 5B. We can expect them to serve similar purposes.

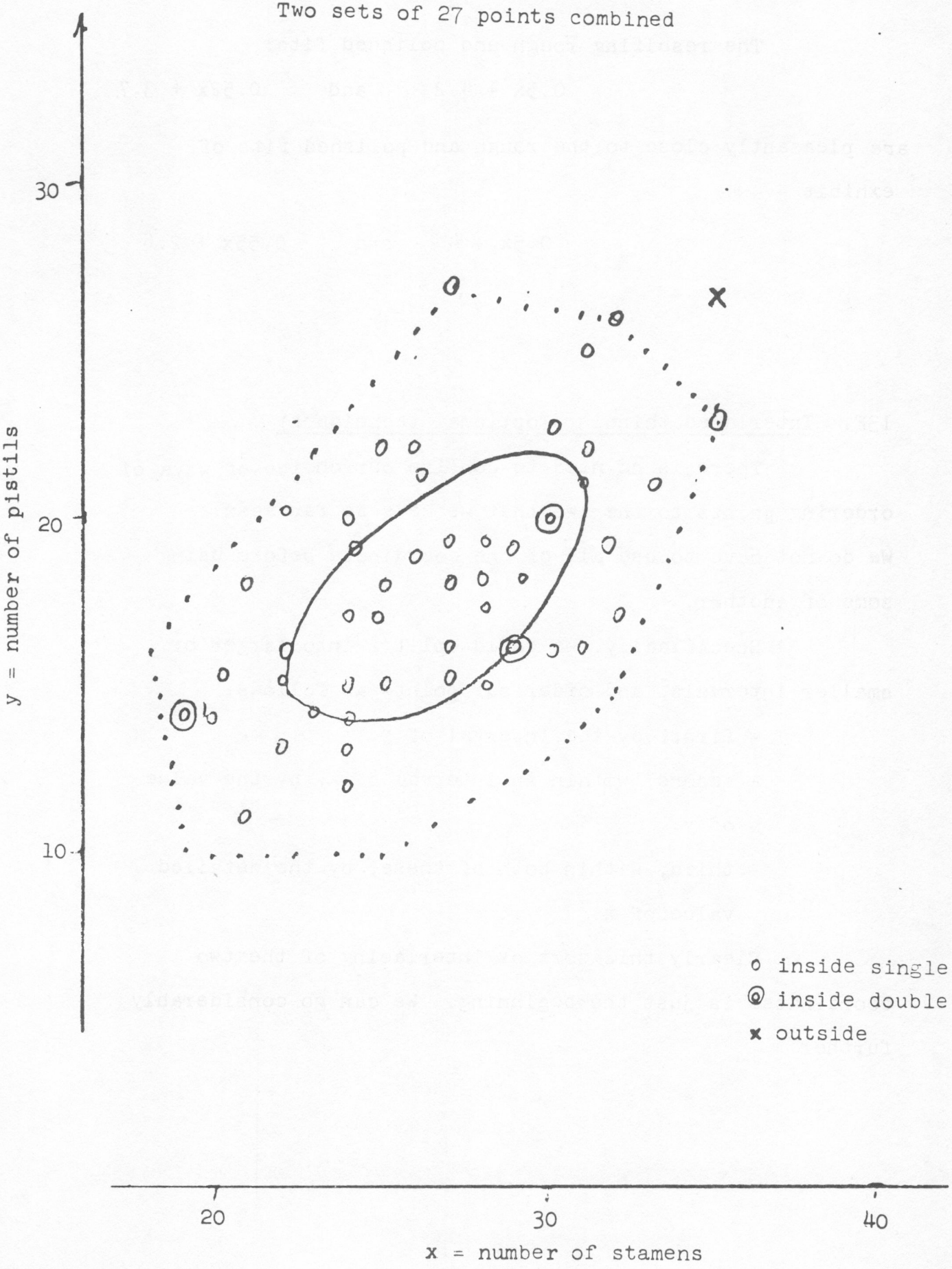
13E. <u>A thin trial tilt</u>

Another way to use a thinned-out set of values is as a basis for calculating a trial tilt. Exhibit 17 puts the first set of 27 points to this use.

exhibit 16 of chapter 13

Combined approach with thinning

Two sets of 27 points combined



The resulting rough and polished fits:

$$0.5x + 4.2 \quad \text{and} \quad 0.52x + 3.7$$

are pleasantly close to the rough and polished fits of exhibit 4

$$0.5x + 4 \quad \text{and} \quad 0.55x + 2.8$$

13F. Interlaced thinning (optional techniques)

There is no need to confine our choice of ways of ordering points to the two that we have so far seen. We do not have to use all of one coordinate before using some of another.

Specifically, we could split x into larger or smaller intervals, and order our points as follows:

- first, by the interval of x.
- second, within an interval of x, by the value of y.
- third, within both of these, by the detailed value of x.

Clearly this sort of interlacing of the two coordinates is just the beginning. We can go considerably further.

exhibit 17 of chapter 13

Using the first set of 27 points to find a rough tilt

x	y	for y	for x	for y	0.5x	Diff	for tilt
19	14	14443	9		9.5	4.5	9
20	14	18555	3H21	14	10	4	3H2.5
21	18	20	5M23	15	10.5	7.5	5M4
22	15		3H24	15	11	4	3H4.5
23	14		tri22.8	14.8	11.5	2.5	tri3.8
24	13				12	1	
24	15				12	3	
24	20				12	8	
25	15				12.5	2.5	
25	17	27595898	9		12.5	4.5	9
26	15	22	3H26	17	13	2	3H4
26	19	27	5M27	18	13	6	5M4.5
26	22		3H27	19	13	9	3H6
27	15		tri26.8	18	13.5	1.5	tri9.8
27	18				13.5	4.5	
27	19				13.5	5.5	
27	27				13.5	13.5	
28	18				14	4	
29	16	169669	9U		14.5	1.5	9
29	19	2013	3H30	16	14.5	4.5	3H1.5
30	16	26	5M31	19	15	1	5M4.5
30	20		3H32	21	15	5	3H5.5
31	16		tri31	18.8	15.5	0.5	tri4.0
31	21				15.5	5.5	
32	19				16	3	
32	26				16	10	
35	23				17.5	5.5	

$$\text{Apparent tilt} = \frac{18.8 - 14.8}{31 - 22.8} = 0.49$$

Trial tilt = 0.5

$$\text{Apparent level} = \frac{3.8 + 4.8 + 4.0}{3} = 4.2$$

Rough fit = 0.5x + 4.2

$$\text{Additional tilt} = \frac{4.0 - 3.8}{31 - 22.8} = 0.02$$

Additional level = 1/3(-0.4 + 0.6 - 0.2) = 0.0

Adjustment to bend for additional tilt = -1/3(.45 + .54 + .62) = -.5

The simple interlace serves for an easy example in exhibit 18. Here the intervals of x are 5 long, and correspond to the lines of the stem-and-leaf displays in exhibit 12. All that we have to do to make the layout of exhibit 18 is to sort these lines out properly and write them down again.

Exhibit 19 shows the 27 selected points with the same two-oval background. The fact that this set may look a little better may well be an accident.

computer version

If we wanted to interlace our numbers more thoroughly, we could alternate digits, thus sorting

(abcde, pqrst)

as if it were

(apbqcrdset)

This is heavily interlaced, but always gives the first coordinate the lead. An interlace very much like that used by some to lace their shoes is a little neater, and possibly a little better. In it

(abcde, pqrst)

is sorted as if it were

(apqbcrsdet)

These patterns work even more neatly on a modern computer where digits are 0 or 1 and many-digit numbers arise easily.

exhibit 18 of chapter 13

Interlaced order (x in 5's, then y in 1's, then x in 1's)

for the 268 points of the Ranunculus data

A The data

x	y	x-detail
15 to 19	10	19
	13	9
	14	9999
	16	8
20 to 24	11	211
	12	00012234
	13	01222234
	14	0002334444
	15	01122223334444444
	16	114
	17	22334
	18	1334444
	19	344
	20	224
	21	2
25 to 29	7	29
	9	9
	10	5
	11	5
	12	55567
	13	589
	14	56677789
	15	5555666677777889
	16	555556667777788889999 9
	17	555566777899
	18	555677777778999
	19	66666777788889
	20	566779
	21	6677
	22	5667779
	23	6799
	24	6799
	27	57
	31	9

exhibit 18 (Cont'd)

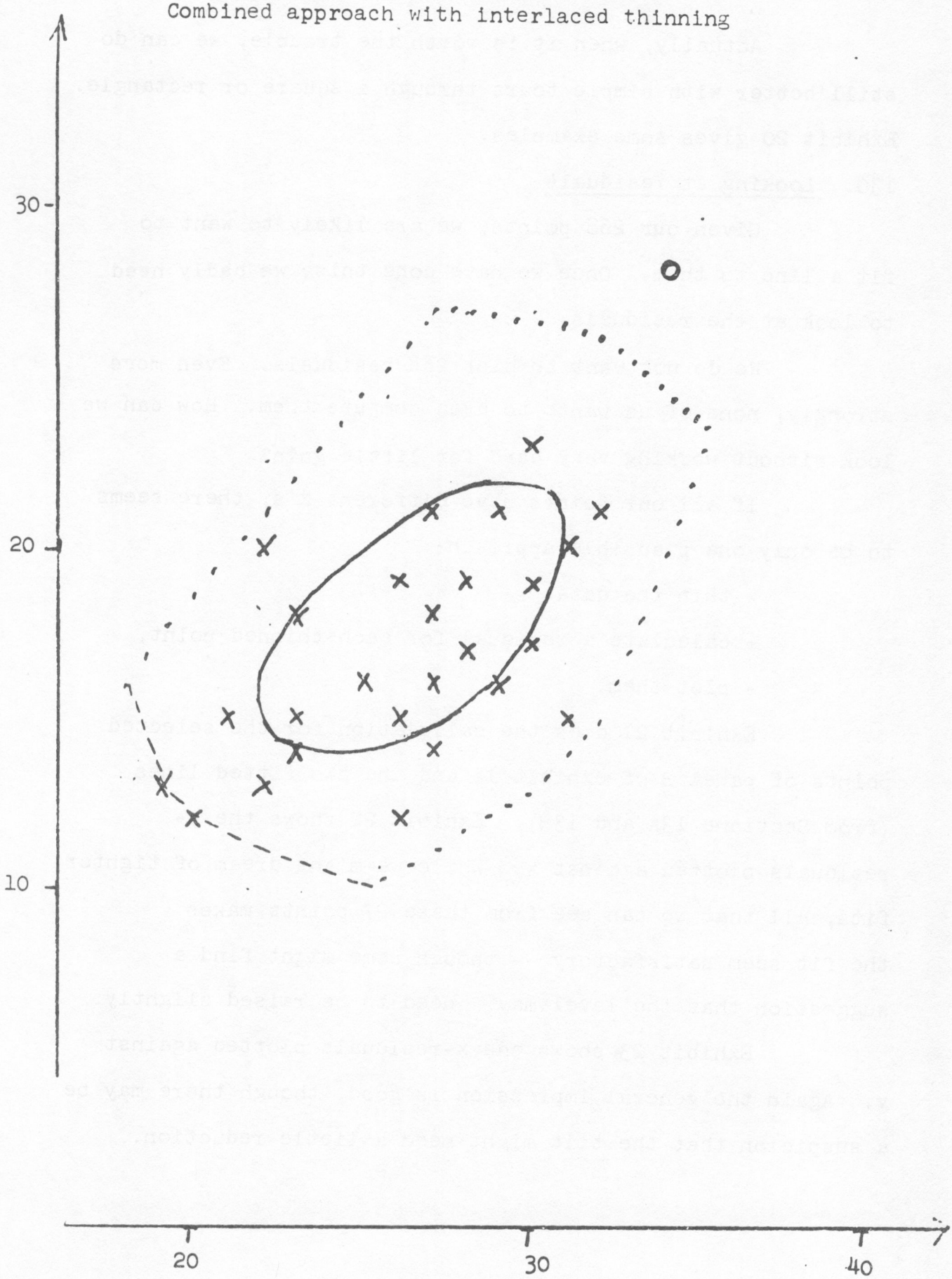
x	y	x-detail
30 to 34	15	312
	16	00011
	17	00000112
	18	00133
	19	0123
	20	00011111223
	21	00111123
	22	00111233
	23	0122
	25	122
	26	022
	28	4
35 to 39	2	37
	23	55
	25	8
	27	55

B) SELECTED POINTS -- in (x,y) form

(19,13), (20,12), (22,13), (24,14), (22,15), (24,15), (23,18), (22,20), (26,12), (27,14), (26,15), (25,16), (27,16), (29,16), (28,17), (27,18), (26,19), (28,19), (27,21), (27,23), (31,15), (30,17), (30,19), (31,20), (32,21), (30,23), (34,28).

exhibit 19 of chapter 13

Combined approach with interlaced thinning



Actually, when it is worth the trouble, we can do still better with simple tours through a square or rectangle. Exhibit 20 gives some examples.

13G. Looking at residuals

Given our 268 points, we are likely to want to fit a line to them. Once we have done this, we badly need to look at the residuals.

We do not want to plot 268 residuals. Even more strongly, none of us wants to even compute them. How can we look without working very hard for little gain?

If all our points have different x's, there seems to be only one plausible approach:

- thin the data,
- calculate a residual for each thinned point,
- plot them.

Exhibit 21 does the calculation for the selected points of panel B of exhibit 12 and the two fitted lines (from Sections 13A and 13B). Exhibit 22 shows the y-residuals plotted against x. While we might dream of tighter fits, all that we can see from these 27 points makes the fit seem satisfactory -- though some might find a suggestion that the level may need to be raised slightly.

Exhibit 23 shows the x-residuals plotted against y. Again the general impression is good, though there may be a suspicion that the tilt might need a little reduction.

exhibit **20** of chapter 13

Simple Tours Through Squares and Rectangles: Some Examples

2	1	16	15
3	4	13	14
6	5	12	11
7	8	9	10

4 × 4

A basic form

7	8	9	10
6	5	12	11
3	4	13	14
2	1	16	15
31	32	17	18
30	29	20	19
27	28	21	22
26	25	24	23

4 × 8

Looks stuck, but replace lower half by

23	22	17	18
24	21	20	19
25	28	29	20
26	27	32	31

and we can go on and on ...

1	4	5	8	9	12	13	14
2	3	6	7	10	11	16	15
31	30	27	26	23	22	17	18
32	29	28	25	24	21	20	19
33	36	37	40	41	44	45	46
34	35	38	39	42	43	48	47
63	62	59	58	55	54	49	50
64	61	60	57	56	53	52	51
65	68	69	72	73	76	77	78
66	67	70	71	74	75	80	79
95	94	91	90	87	86	81	82
96	93	92	89	88	85	84	83

8 × 12

Easily cut to 8×2, 8×4, 8×6, 8×8 or 8×10

Easily extended to 8×14, 8×16, 8×18, 8×20, ---

A modified form works for 2j × 2k for any integers j and k.

We can do 2j × (2k+1) even more easily.

exhibit **21** of chapter 13

Calculation of residuals for 27 selected points

(see panel B of exhibit 12)

x	y	0.5x+3.8	Residual*	0.77y+13.1	Residual**
19	14	13.3	0.7	23.9	-4.9
20	14	13.8	0.2	23.9	-3.9
21	18	14.3	3.7	27.0	-6.0
22	15	14.8	-0.2	24.6	-2.6
23	14	15.3	-1.3	23.9	-0.9
24	13	15.8	-2.8	23.1	0.9
24	15	15.8	-0.8	24.6	-0.6
24	20	15.8	+4.2	28.5	-4.5
25	15	16.3	-1.3	24.6	+0.4
25	17	16.3	0.7	26.2	-1.2
26	15	16.8	-1.8	24.6	1.4
26	19	16.8	2.2	27.7	-1.7
26	22	16.8	5.2	30.0	-4.0
27	15	17.3	-2.3	24.6	2.4
27	18	17.3	0.7	27.0	0.0
27	19	17.3	1.7	27.7	-0.7
27	27	17.3	9.7	33.9	-6.9
28	18	17.8	0.2	27.0	1.0
29	16	18.3	-2.3	25.4	3.6
29	19	18.3	0.7	27.7	1.3
30	16	18.8	-2.8	25.4	4.6
30	20	18.8	1.2	28.5	1.5
31	16	19.3	-3.3	25.4	5.6
31	21	19.3	1.7	29.3	1.7
32	19	19.8	-0.8	27.7	4.3
32	26	19.8	6.2	33.1	-1.1
35	23	21.3	1.7	30.8	4.2

Residual* = y - 0.5x - 3.8.

Residual** = x - 0.77y - 13.1.

exhibit **22** of chapter 13

27 thinned y-residuals

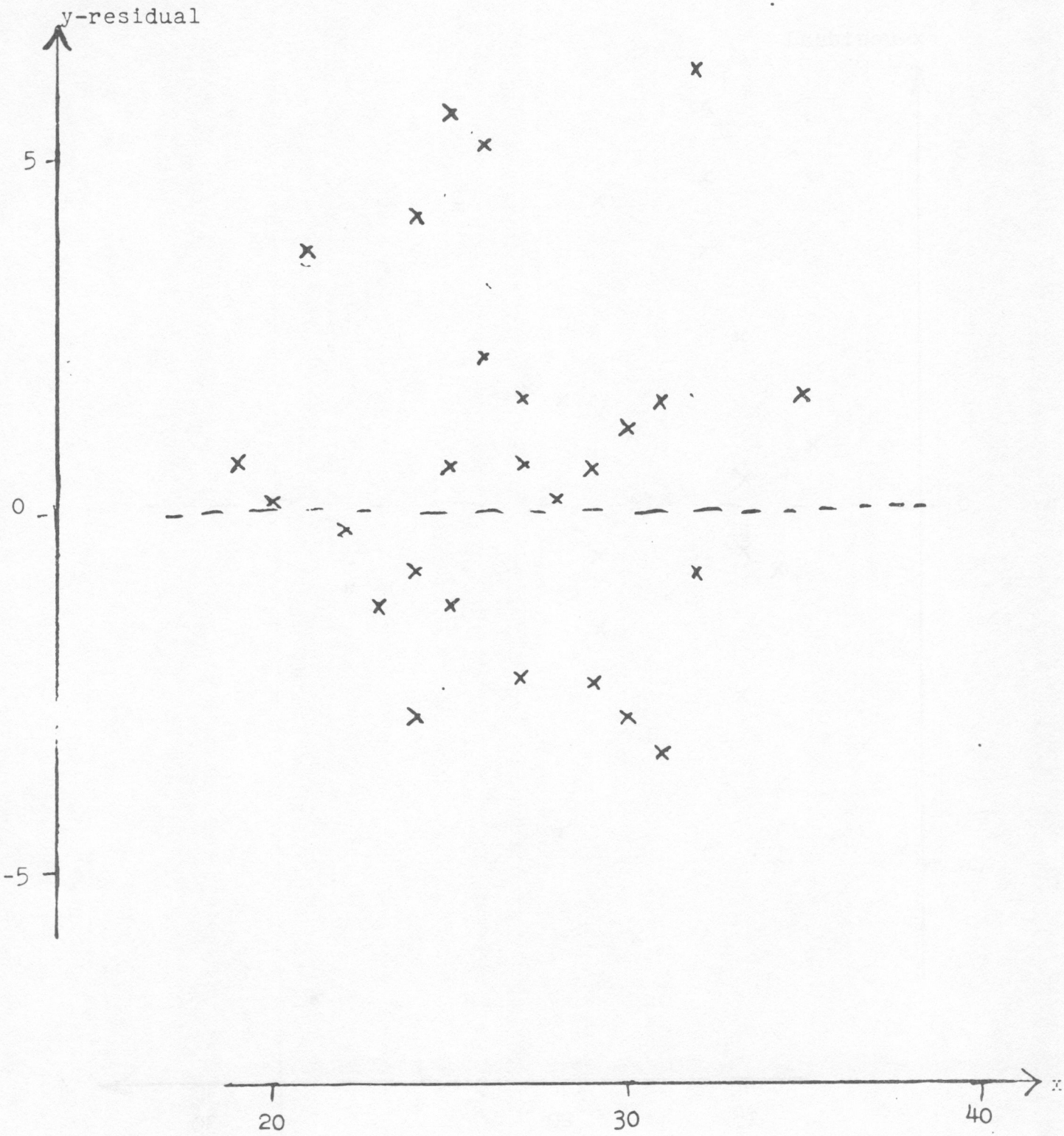
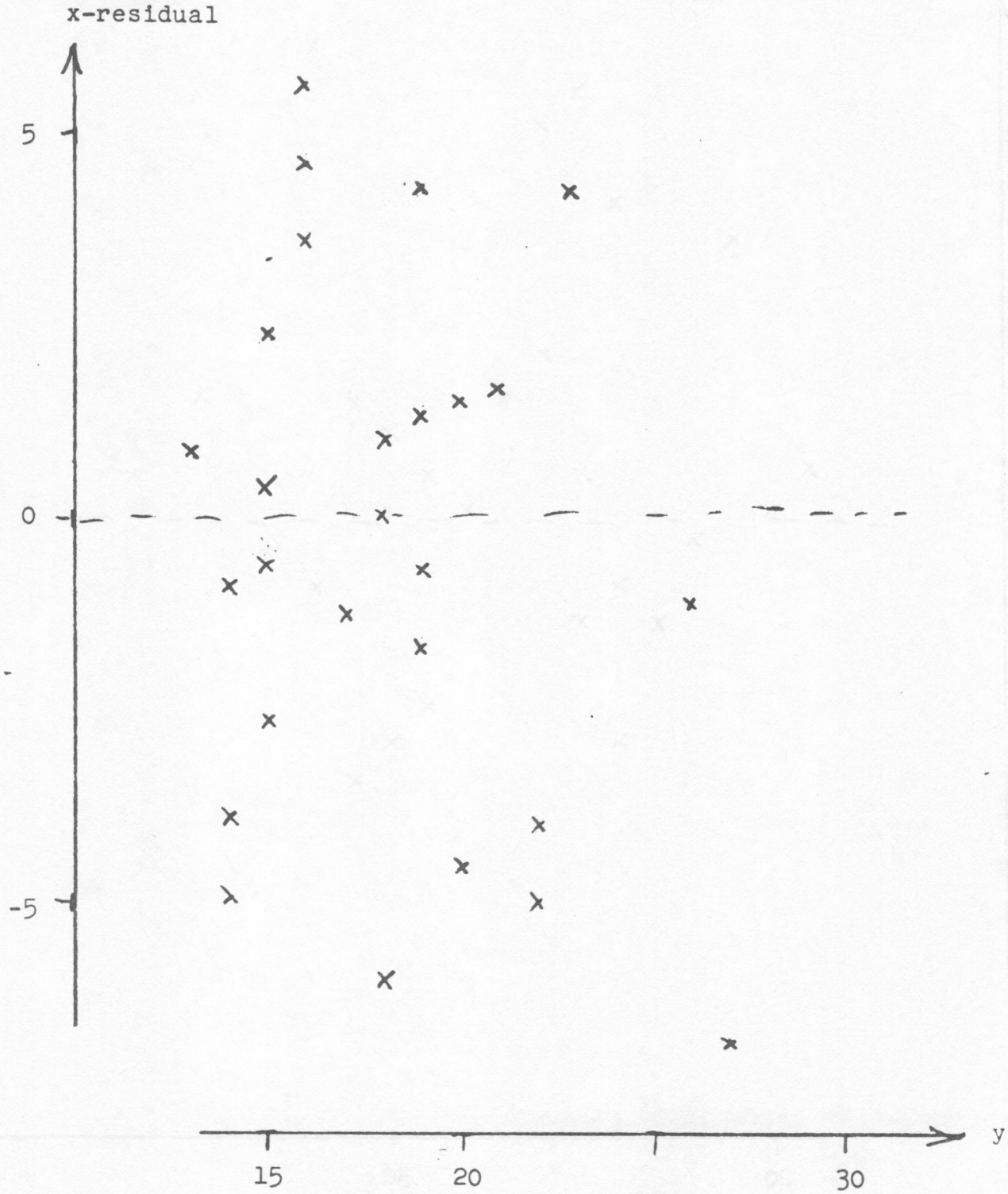


exhibit 23 of chapter 13

27 thinned x-residuals



The moral here is simple -- and vital:

- it is far, far better to have looked at a carefully selected part of the data than never to have looked at all. It may seem honorable and high-principled to say that you will only look at all the data, but if that means you don't look at even a part, you have cut yourself off from all chance of learning.

* working with x-arrays *

Sometimes we don't have to reduce ourselves to thinning the data in order to look at residuals. If we have many points with the same x-value, we can treat these points as a batch, summarize each batch, and then adjust the batch summaries for a fit. This is, of course, the situation we enjoy in our example.

For x = 27, exhibit 1 shows that there are 38 y's. Their 5-number summary is

38 *HMH* = 12, 16, 18, 20, 27

For x = 27, the fitted value is .55x + 2.8 = 17.6. Since this same value is fitted to all 38 values in this batch, we can get the 5 numbers summarizing the residuals by subtracting 17.6 from each of the 5 numbers summarizing y. This means that the residuals have

38 *HMH* = -5.6, -1.6, 0.4, 2.4, 9.4

Exhibit 24 shows the numbers that come from doing this for each x.

exhibit 24 of chapter 13

The *HMH* summaries for the batches formed of all residuals corresponding to a given x.

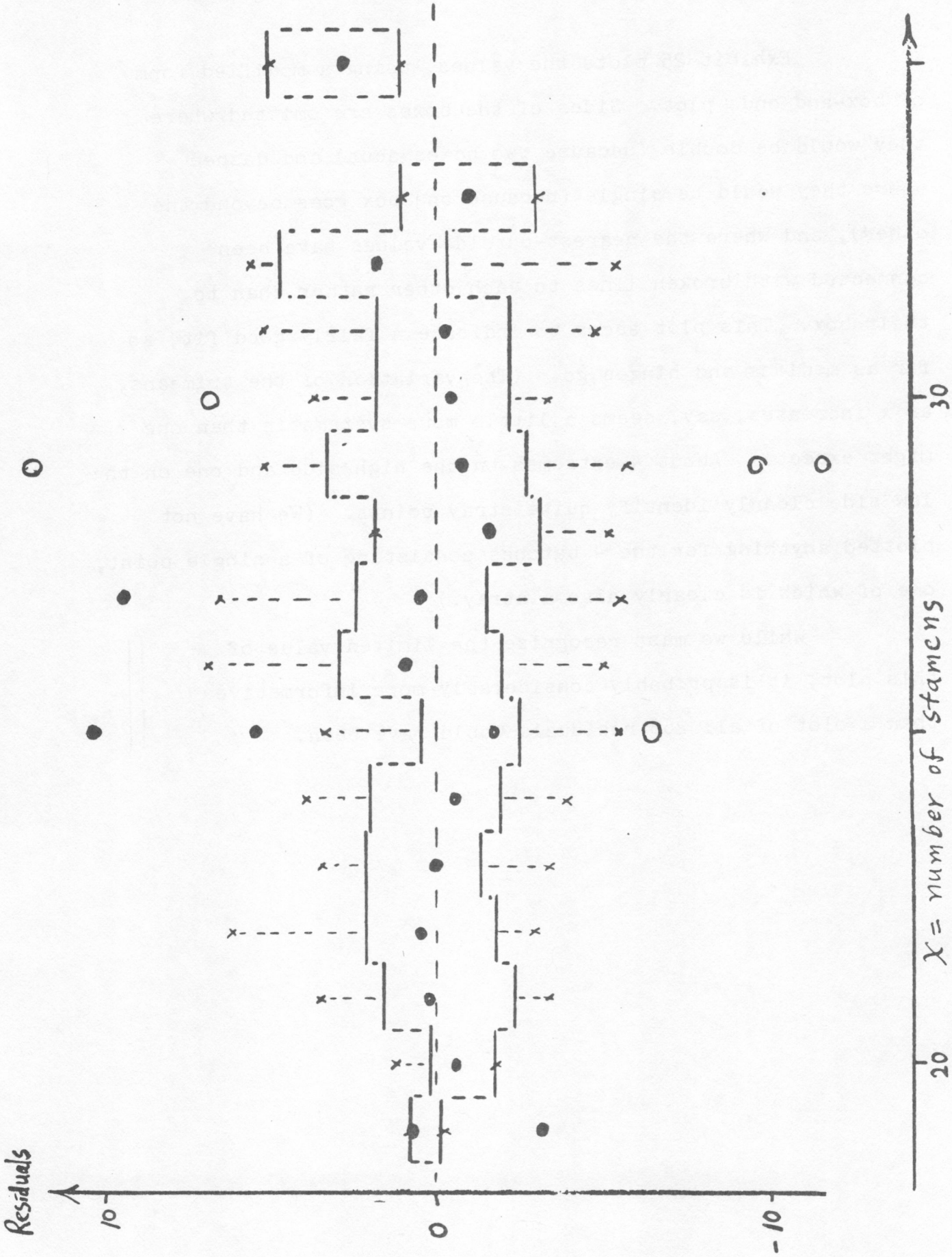
(#)	x	*	H	M	H	*	Trimean
(1)	18	-	-	(3.3)	-	-	(3.3)
(6)	19	-3.2	-.2	.8	.8	.8	.6
(8)	20	-1.8	-1.8	-.3	.2	1.2	-.6
(9)	21	-3.4	-2.4	.6	1.6	3.6	.1
(16)	22	-2.9	-1.9	.1	2.1	6.1	.4
(12)	23	-3.4	-1.4	--.4	2.1	3.6	.0
(22)	24	-4.0	-2.0	-1.0	2.0	4.0	-.5
(26)	25	-6.6	-2.6	-.5	.4	10.4	-.8
(26)	26	-5.1	-2.1	1.4	2.9	6.9	.9
(38)	27	-5.6	-1.6	.4	2.4	9.4	.4
(14)	28	-5.2	-3.2	-2.2	.8	.8	-1.7
(23)	29	-11.8	-2.8	-1.8	3.2	12.2	-.8
(20)	30	-3.3	-2.3	-.9	1.7	6.7	-.5
(20)	31	-4.8	-2.3	.2	1.7	5.2	-.3
(13)	32	-5.4	-.4	1.6	4.6	5.6	1.8
(7)	33	-3.0	-3.0	-1.0	1.0	1.0	1.0
(1)	34			(6.5)			(6.5)
(4)	35	1.0	1.0	3.0	5.0	5.0	2.8
-	36						
(1)	37			(-21.2)			(-21.2)
(1)	38			(1.3)			(1.3)

Exhibit 25 plots the values, using a modified form of box-and-ends plot. Sides of the boxes are omitted where they would be double (because two boxes abut) and dashed where they would be single (because one box goes beyond the other), and where the nearest-outside values have been connected with broken lines to each other rather than to their box. This plot seems to indicate a fairly good fit, so far as medians and hinges go. (The variation of the trimeans, as x increases, say, seems a little more systematic than one might expect.) About 4 extremes on the high side and one on the low side clearly identify quite stray points. (We have not plotted anything for the 4 batches consisting of a single point, one of which is clearly also a stray.)

While we must recognize the limited value of this plot, it is probably considerably more informative than a plot of all 268 residuals would have been.

exhibit 25 of chapter 13

The summaries of the batches of residuals for each x



*****

13H. How far have we come?

The purpose of this chapter was to learn how to avoid being drowned by moderately large amounts of (x,y) data.

We learned how to use a two-way table of counts as a basic form of giving all the data. Trimeans -- or medians for few y's -- for each x-array led rather easily to a fitted line.

The same table and its modifications led to the basic information for a schematic plot.

To plot individual points against a schematic-plot background -- or to plot residuals -- we found it wise to thin the data by first arranging the point in a systematic order and then taking every kth point. We use as good a choice of order as we know -- or can afford.

Another way to look at residuals, probably better, is to make a schematic plot of each x-array. (If x's tend to be all different, we would make such a plot for groups of points with nearly equal x's.)

*****

13K. Using spread-out values (optional technique)

Turning back to exhibit 7, we see that we are facing more irregularity than seems essential. The lower eighth for x+y would change a whole unit if one low point were set aside. This is because rank 34 would then move from 35 to 36, were there 7 values at 36 and 12 values at 37.

Had we used spread-out values, rank 34 would have been assigned to 36.43 before the change and to 36.54 afterward. The resulting shift of 0.11 seems much more reasonable. Exhibit 26 repeats the analysis of exhibit 7 using spread-out values. The results are often only slightly different, but the low eighth for y is moved 0.48 -- almost the greatest amount possible. As a result the low T-point is moved a little more than a whole unit.

In the calculations of exhibit 26, we found reason to spread out the values that fix eighths and hinges, but not to spread out the values being examined as to whether or not they are outside. Indeed, since we had only to deal with single occurrences of each value, spreading out would have made no difference.

Spreading out values that may be outside can make a difference as panel F of exhibit 27 will show. This exhibit displays, subject to the reader's completion, the coordinatewise skipping analyses for eighths and hinges, using raw values and using spread-out values. Coordinatewise skipping proceeds normally, in each case, following, however, different paths to different results.

exhibit 26 of chapter 13

Calculation of eights and hinges for x, y, x+y, x-y

using spread-out values

A) BREAKS, from exhibit 7, and HINGES and EIGHTHS for SPREAD-OUT VALUES

x	**depth**	**spread x**	**x**	**depth**	**spread x**
21h	25h		31h	27h	
22h	41h		30h	47h	
23h	53h		29h	67h	
24h	75h		28h	90h	
	34	22.03=21h+(8h/16)		34	31.17=30h+(13h/20)
	67h	24.14=23h+(14/22)		67h	29.50=29h

y	**depth**	**spread y**	**y**	**depth**	**spread y**
13h	33h		22h	27h	
14h	55h		21h	42h	
15h	70h		20h	55h	
	34	13.52=13h+(h/22)	19h	74h	
	67h	14.84=14h+(12/35)		34	22.07=21h+(8h/15)
				67h	19.87=19h+(7/19)

x+y	**depth**	**spread x+y**	**x+y**	**depth**	**spread x+y**
35h	27h		52h	29h	
36h	34h		51h	41h	
37h	46h		50h	52h	
38h	56h		49h	58h	
39h	69h		48h	66h	
	34		47h	75h	
	67h			34	
				67h	

x-y	**depth**	**spread x-y**	**x-y**	**depth**	**spread x-y**

Problem: Complete this panel.

exhibit 26 (cont'd.)

B) COMPARISON of RESULTS -- raw, then spread out

	for x		for y		for x+y		for x-y	
T	13	12.89	6	4.97	20	-	-	-
S	16h	16.67	9	8.49	27	-	-	-
E	22	22.03	14	13.52	36	-	-	-
H	24	24.14	15	14.84	39	-	-	-
H	29h	29.50	20	19.87	48	-	-	-
E	31	31.17	22	22.07	52	-	-	-
S	36h	36.53	27	27.10	61	-	-	-
T	40	40.31	30	30.62	68	-	-	-
dh	5h	5.36	5	5.03	9	-	-	-
dl	9	9.14	8	8.55	16	-	-	-

changes, unit = 0.01, signs neglected

	(8 values)	(median)
for T:	11,31,103,62,--,--,--,--,	--
for S:	17, 3, 51,10,--,--,--,--,	--
for E:	3,17, 48, 7,--,--,--,--,	--
for H:	14, 0, 16,13,--,--,--,--,	--

Problem: Complete this panel.

exhibit 27 of chapter 13

Combining spread-out and coordinatewise skipping in calculating eighths and hinges, and a further alternative

A) COORDINATEWISE SKIPPING CALCULATIONS for x

raw

		34	**0**	**34**	**0**	
	T	**67h**	13	**67**	13	
	S	16h		17		
	E	22		22		
268	H	24		24		stops
	H	29h		29		
	E	31		31		
	S	36h		36		
	T	**67h**	40	**69**	40	
		34	**2**	**36**	**2**	
	dh	5h		5		
	dl	9		9		

spread-out

		34	**0**	**34**	**0**	
	T	**67h**	10.89	**67**	12.98	
	S	16.67		16.58		
	E	22.03		22.03		
268	H	24.14		24.11		**56**
	H	29.50		29.56*		
	E	31.17		31.08		
	S	36.53		36.01		
	T	**67h**	40.31	**69**	40.13	
		34	**2**	**36**	**2**	
	dh	5.36		5.45		
	dl	9.14		9.05		

B) COORDINATEWISE SKIPPING CALCULATIONS for y

raw

		34	**2**	**35h**	**2**	**35**	**2**	
	T	**67h**	4.97	**68h**	5.21	**67h**	5.41	
	S	8.49		8.64		8.76		
	E	13.42		13.59		13.57		
268	H	14.84		14.87		14.84		
	H	19.87		19.82		19.65		stops
	E	32.07		21.97		21.73		
	S	27.10		26.92		26.54		
	T	**67h**	30.62	**68h**	30.35	**71h**	29.89	
		34	**2**	**35h**	**6***	**39**	**6**	
	dh	5.03		4.95		4.81		
	dl	8.55		8.38		8.16		

* We shall return to this count.

Problems: Complete this panel.

C and D) COORDINATEWISE SKIPPING CALCULATIONS for **x+y** and **x-y**

Problem: Complete this panel

exhibit 27 (cont'd.)

E) A REMARK and an ALTERNATIVE

In panel B, when the high S for y fell from 27.10 to 26.92 we increased the high-outside count from 2 to 6. The y's involved are 31,28,27,27,27,27. The question is: Should we spread out the 27's? If we did, they would fall according to

x	high depth	spread
27h	2h	
26h	6h	
	3	27 - 38
	4	27 - 12
	5	26 - 88
	6	26 - 62

This would change our high-outside count, putting 2 more above 27.10 and 2 fewer above 26.92. The result would be:

		34	**2**		**35**	**2**		**35**	**2**	
	T	**67h**		4.97	**68**		5.27	**68**		5.34
	S		8.49			8.69			8.74	
	E	13.52			13.57			13.57		
268	H	14.84			14.86			14.86		stops
	H	19.87			19.74			19.69		
	E	22.07			21.87			21.80		
	S		27.10			26.75			26.63	
	T	**67h**		30.62	**70**		30.17	**71**		30.03
		34	**4**		**37**	**5**		**88**	**5**	
	dh	5.03			4.88			4.83		
	dl	8.55			8.30			8.23		

Problems:

(1) Apply the alternative, where relevant, to panels A, C, or D.

(2) Discuss the advantages and disadvantages of the alternative. How much is it likely to matter?

13L. Still more points? (optional background)

We have learned to cope fairly well with 268 (x,y) pairs. What if there had been 2500? Or 25,000? Or triples?

* more variables *

We are only learning about exploring (x,y) pairs in this account. What if we had (x,y,z) triples -- or quadruples -- or more?

So long as we can have a two-way table of counts, we can find trimeans of x-arrays and fit lines to them.

If all the x's are different, we can use x-arrays for intervals of x rather than individual values. It may pay to make a first fit based on a thinning, only collecting points with adjacent x values after we have roughly adjusted y values.

To make schematic plots, our only problem is to find eighths and hinges. We can use thinned values whenever this will do well enough.

To use a combined approach, showing some individual points, may be more difficult. With 2500 points, we would be reluctant to thin to 25 -- usually, we may suppose, rightly so. For many purposes thinning to 100 points might be workable. (And if we have labored to collect 2500 points, the arithmetic of working with 100 is not likely to seem large and out of proportion.)

(Partly since it is hard to plot in space -- and harder still to plot in 4 or more dimensions -- much of what is done with triple, quadruple -- data is done two variables at a time. Similar techniques will apply.)

* even more points *

As we go to, say, 25,000 points, something has to give. Thinning down to few enough points to make a plot we can look at is keeping just too small a fraction of the data. (At least it would be if we had any real need for that many points.) Something new needs to be added. We must plan to plot composite points, something that would already have been likely to help us with, say 2500 points.

There seem to be only a few natural reasons for wanting 25,000 points:

- we are interested in comparisons between one part and another, not in all 25,000 together.
- we are interested in certain rare events, and need to look at all points of -- or resembling -- certain sorts -- but we only need a small fraction of the rest.

- we are trying to look at small differences to be revealed after a fitting process (often a process involving more than just an x and a y).

If we adjust our techniques to match our real questions, neither of the first two reasons will lead to our being drowned with data. The third reason calls for more thought.

Suppose we have 25,000 sets of data, perhaps involving 5 variables. (Clearly we must be using a computer.) If we are willing to be a little optimistic -- hoping that the kinds of fitting that prove to be needed will behave rather like fitting straight lines -- we can take a first cut by thinning the data to 1000 sets.

We can use interlace on 5 variables with little extra difficulty, especially binary interlace on a computer. The following steps can now be taken:

- order the 25,000 sets on the basis of interlace.
- break up into 1000 blocks, each of about 25 adjacent sets -- we should adjust the breaks to match large jumps in values corresponding to breaks in the binary interlace.
- replace each block by a single newly-defined set, each variable in the new set being the trimean of the 25 values of that variable in the 25 sets of the block.
- work with the 1000 summarized values until we either learn of the need of re-expression or have found

a way to concentrate the meaning that concerns us into the values of quite small residuals.

- once either of these situations is found, go back to the 25,000 re-expressing, re-ordering, re-blocking, and forming new trimeans.

It is a rare set of data that is at all manageable but cannot be handled by such an approach.

*****

13P. Problems

(1) Starting with the 2nd, rather than the 4th, point in the list of panel A of exhibit 12, select every 10th point. Plot the selected points and compare with exhibit 13.

(2) Do the same with your choice of starting point among the first 10 in the list of exhibit 12.

(3) Plot every 5th point, beginning with the 2nd, in the list of panel A of exhibit 12. Comment on how this compares with the result of problem 1.

(4) Production as a percentage of a smooth fit for 204 months; industrial production and pig-iron production (see exhibit 28). Thin to 20 or 21 points in five ways (by using spacing 10 with 5 different starts. Plot each thinned batch on a separate sh-et of tracing paper. How satisfied are you with the agreement of the five sets?

(5) 500 double tosses of 12 coins, only 4 being tossed anew (see exhibit 29). Find both sets of array trimeans and plot on the same sheet. Find the anchored line, and plot on another sheet with the array trimeans.

(6) 1336 pairs of adjacent vehicles on a highway (see exhibit 30). Thin to 34 points (by spacing 40). Plot the thinned batch.

(7) For the same data (exhibit 30), calculate array trimeans for distances at each given speed, and plot.

exhibit 28 of chapter 13

Pig-iron production and industrial production
as percentages of trends
(monthly, 1897-1913)

Industrial production as % of trend	Pig-iron production as % of trend								Total
	55	65	75	85	95	105	115	125	
125								15	15
115						6	34	1	41
105					5	51	6		62
95				3	33	1			37
85			2	24	3				29
75			7	2					9
65		2	1						3
55	6	2							8
Total	6	4	10	41	41	58	40	16	204

Source: E. Harold T. Davis and W. F. C. Nelson 1935.
Elements of Statistics with Applications to Economic Data.
Principia Press. Table A on page 258.

exhibit 29 of chapter 13

500 double tosses of 12 coins displayed in two ways
(only 4 coins retossed)

A) FULL PATTERN

Heads on 2nd toss	Heads on first toss											Total
	2	3	4	5	6	7	8	9	10	11	12	
11							1					1
10						1	3	3	3			10
9				1	1	4	13	5	5	3	1	33
8				1	8	19	13	10	1			52
7		1	1	15	29	27	13	7	1			94
6		4	12	30	36	19	10	7	1			119
5		4	21	30	26	13	4					98
4	3	1	12	20	9	5	2					52
3	2	8	9	7	4							30
2	4	2	3	2								11
Total	9	20	58	106	113	88	59	32	11	3	1	500

B) CONDENSED PATTERN with SUPPLEMENTARY DETAIL

2nd Toss	Heads on first toss							Total
	lo	4	5	6	7	8	hi	
hi	.	.	.	.	1	4	6	11
9	.	.	1	1	4	13	14	33
8	.	.	1	8	19	13	11	52
7	1	1	15	29	27	13	8	94
6	4	12	30	36	19	10	8	119
5	4	21	30	26	13	4	.	98
4	4	12	20	9	5	2	.	52
lo	16	12	9	4	.	.	.	41
Total	29	58	106	113	88	59	47	500

----Supplementary detail; arranged (1st toss, 2nd toss)----

Once each: (3,7), (3,4), (7.10), (8,11), (10,8), (10,7), (10,6), (12,9). Twice each: (2,3), (3,2), (5,2). Three times: (2.4), (4,2), (8,10), (9,10), (10,10), (11,9). Four times: (2,2), (3,6), (3,5), (6,3). Five times: (9,9), (10,9). Seven times: (5,3), (9,7), (9,6). Eight times: (3,3). Nine times: (4,3).

exhibit 29 (cont'd.)

Ten times: (9,8). (Check: (8)(1) + (3)(2) + (6)(3) + (4)(4) + (5)(2) + (7)(3) + (8)(1) + (9)(1) + (10)(1) = 106 here. In condensed table, hi or lo = 29 + 47 + 41 + 11 = I28, from which must be subtracted 16 + 0 + 6 + 0 = 22 (to avoid double counting), giving 128-22 = 106. Next 106=106, check.)

Source: R. Pearl 1923. Medical Biometry and Statistics, Philadelphia, page 298. Adapted from A. D. Darbishire 1907. Some tables for illustrating statistical correlation. Mem. and Proc. Manchester Lit. and Phil. Soc. 51 (20 pages). (Used by H. T. Davis and W. F. C. Nelson 1935. Elements of Statistics with Applications to Economic Data. Principia Press.)

exhibit 30 of chapter 13

1336 separations of adjacent vehicles on a highway
(center-to-center distances in 4-foot steps*,
speeds in 2 m.p.h. steps**)

----------long distance----------

Distance	Distribution over speed
140h:	1 at 29, 1 at 35
136h:	1 at 29, 1 at 33, 1 at 35
132h:	1 at 29, 1 at 35
128h:	none
124h:	1 at 35, 1 at 47
120h:	1 at 25
116h:	1 at 25, 1 at 31, 1 at 33
112h:	1 at 17, 1 at 23, 1 at 27, 1 at 37, 1 at 39
108h:	1 at 19, 2 at 23, 1 at 27, 2 at 31, 1 at 47

----------high speeds----------

Speeds	Distribution over distance
53:	1 at 36h
51:	1 at 36h, 1 at 60h
49:	1 at 40h, 1 at 44h, 1 at 48h
47:	1 at 28h, 1 at 36h, 1 at 52h, 1 at 60h,(1 at 108h),(1 at 124h)
45:	1 at 36h, 1 at 40h, 1 at 44h, 1 at 56h, 2 at 60h, 2 at 80h

----------short distances----------

Distance	Distribution over speeds
16h:	6 at 1, 2 at 3, 2 at 7

----------low speeds----------

Speeds	Distribution over distance
1:	(6 at 16h), 23 at 20h, 19 at 24h, 4 at 28h, 1 at 36h, 2 at 60h
3:	(2 at 16h), 13 at 20h, 24 at 24h, 22 at 28h, 6 at 32h, 4 at 36h, 3 at 40h, 1 at 49h

(cont'd)

exhibit 30 (cont'd)

center of pattern

Distance in feet	speed in miles per hour																			
	5	7	9	11	13	15	17	19	21	23	25	27	29	31	33	35	37	39	41	43
104h							3	.	.	1	1	1	1		2	1	.	.	.	1
100h					1	.	.	.	.	.	1	.	.		.	.	.	2	.	1
96h									2	3	1	.	2	1	2	.	1	2	1	2
92h									1	1	4	.	1	.	2	.	1	.	.	
88h						1	.	.	.	4	5	4	1	5	1	.	.	1	1	
84h							.	1	1	3	3	2	1	5	2	2	3	.	.	
80h			1	.	.	1	1	1	1	.	1	2	2	1	2	3	.	2	.	.
76h						1	.	2	1	.	2	.	3	3	5	4	1	1	.	1
72h						.	.	.	1	.	3	.	5	5	1	7	3	4	1	2
68h					1	.	.	1	2	4	2	2	1	3	4	3	5	4	1	2
64h			2	.	.	1	3	1	2	3	1	1	11	3	12	2	6	3	2	1
60h			1	.	.	1	.	1	1	2	3	.	6	10	9	6	2	4	1	.
56h		1		3	2		1	1	3	4	2	.	9	9	11	5	4	4	1	.
52h	2	1	2	1	1	.	2	3	3	2	9	.	8	6	9	6	2	3	.	.
48h	3	4	2	.	2	2	1	3	3	4	4	3	17	6	9	4	6	7	.	1
44h	1		6	2	6	4	7	1	3	1	.	.	14	8	11	11	3	4	.	.
40h	2	7	13	9	9	3	6	2	6	6	7	2	19	8	8	6	2	3	3	.
36h	12	9	14	2	5	5	3	2	3	7	11	14	5	3	7	10	2	1	2	.
32h	6	9	6	6	11	3	5	5	11	2	6	6	10	9	5	5	3	3	1	.
28h	26	17	8	7	8	6	1	8	6	8	6	1	10	5	2	5	1		1	.
24h	19	17	4	2	2	5	6	3	1	4	5	5	3	.	1	2	.	.	1	
20h	3	3	4	3	2	.	1	1	.	1	2	.	.	1	.					

* Thus 124h covers 123 to 126, while 128h covers 127 to 130

** Thus 29 covers 28.0 to 29.9, while 31 covers 30.0 to 31.9.

Source: B.D. Greenshields and F. M. Werda 1952. Statistics with applications to Highway Traffic Analysis. Ens. Foundation for Highway Traffic Control. Saugatuck, Connecticut. Table III.4 between pages 114 and 115.

(8) For the same data (exhibit 30), calculate array trimeans for speeds at each given distance, and plot.

(9) For the same data (exhibit 30), calculate trimeans for 1/distance at each given speed and plot. Do you prefer this plot to that of problem 5?

(10) For the same data (exhibit 30), calculate trimeans for 1/speed at each given distance and plot. Do you prefer this plot to that of problem 6?

(11) How nearly can you find re-expressions of one or both of distance and speed to make both array trimean plots straight lines? (The same re=expression is to be used when fixing an array as when reporting a trimean.)

(12) Plot both of these near straight lines and the 39 points of problem 6 on the same plot.

(13) 1370 treated telephone poles; depth of sapwood and depth of penetration of treatment (see exhibit 31). Thin to 27 or 28 points (by spacing 50). Plot the thinned batch. Fit both lines (x vs. y and y vs. x) to the trimmed points. Plot lines and thinned points.

(14) Breadths of 1532 pairs of eggs from the same nest of the Common Tern (see exhibit 32). Cut down to a batch of 30 or 31 (by spacing 50). Plot the resulting batch. Fit 2 lines to the 30 or 31 points. Plot these lines and the 30 or 31 points on a second sheet.

exhibit 31 of chapter 13

Depths of sapwood and treatment penetration
in 1370 telephone poles

Penetration (inches)	Depth of sapwood*												Total
	1.0	1.3	1.6	1.9	2.2	2.5	2.8	3.1	3.4	3.7	4.0	hi	
hi	.	.	.	.	.	.	.	1	4	8	8	25	46
2.8	.	.	.	.	.	.	2	7	12	7	14	9	51
2.5	.	.	.	.	.	2	12	6	11	11	12	11	65
2.2	.	.	.	.	2	7	18	24	27	28	15	9	131
1.9	.	.	.	2	10	19	22	36	22	24	9	11	155
1.6	.	.	5	14	39	34	45	42	28	21	10	15	253
1.3	.	1	11	36	48	59	51	40	29	13	10	7	305
1.0	1	12	33	41	42	50	37	22	15	10	2	6	271
.7	1	15	11	13	11	14	6	10	3	1	2	1	88
.4	.	1	2	.	1	1	.	.	.	.	.	.	5
Total	2	29	62	106	153	186	193	188	151	123	82	95	1370

*center values in inches.

supplementary detail for "hi" cases
arranged (sapwood, penetration)

Once each: (3.1,3.1), (4.0,3.7), (4.3,3.7), (4.6,3.7), (4.6,2.5), (4.6,0.7), (4.9,4.3), (4.9,3.1), (4.9,2.2), (4.9.1.9), (4.9,1.0), (5.2,4.6), (5.2,4.0), (5.2,3.7), (5.2,3.1), (5.2,1.0), (5.5,2.5). Twice each: (3.4,3.4), (3.4,3.1), (4.0,3.1), (4.6,4.0), (4.6,3.4), (4.9,3.7), (4.9,2.8), (4.9,2.4). Three times: (4.3,3.4), (4.6,3.1), (4.6,2.8), (4.6,2.2), (4.6,1.9), (4.6,1.3), (4.9,1.6). Four times: (3.7,3.4), (3.7,3.1), (4.3,2.8), (4.3,1.3), (4.3,1.0). Five times: (4.0,3.4), (4.3,3.1). Six times: (4.3,2.2). Seven times: (4.3,2.5), (4.3,1.9), (4.3,1.6).

Source: W. A. Shewhart 1931. Economic Control of the Quality of Manufactured Product. D. Van Nostrand. Figure 27 on page 69.

Exhibit 32 of chapter 13

Egg breadths of pairs of eggs from the same nest
(Each pair entered twice)

Breadth of second egg*	Breadth of first egg (in centimeters)*									Total
	lo	2.92	2.97	3.02	3.07	3.12	3.17	3.22	3.27	
lo	14	9	8	3	.	4	.	.	.	38
2.92	9	20	22	25	9	2	.	.	.	87
2.97	8	22	38	69	29	13	2	.	.	181
3.02	3	25	69	104	89	70	10	5	.	375
3.07	.	9	29	89	72	80	22	6	.	307
3.12	4	2	13	70	80	112	51	14	2	348
3.17	.	.	2	10	22	51	22	14	5	126
3.22	.	.	.	5	6	14	14	6	7	52
3.27	.	.	.	.	.	2	5	7	4	18
Total	38	87	181	375	307	348	126	52	18	1532

*Cell mid-values. 292 stands for at least 2.90 but not as much as 2.95, etc.

Supplementary detail for "lo" and "hi" cases
(each pair still entered twice)

Once each: (2.97,2.72), (3.02,2.72), (2.82,2.77), (3.12,2.77), (2.77,2.82), (2.97,2.82(, (3.02,2.82), (3.02,2.87), (3.12,2.87), (2.72,2.97), (2.82,2.97), (2.72,3.02), (2.82,3.02), (2.87,3.02), (2.77,3.12), (2.87,3.12). Twice each: (2.72,2.67), (2.67,2.72), (2.72,2.72), (2.92,2.82), (3.12,2.82), (2.82,2.92), (2.82,3.12). Three times: (2.87,2.82), (2.82,2.87). Six times: (2.97,2.87), (2.87,2.97). Seven times: (2.92,2.87), (2.87,2.92). (62=62,check).

Source: D. M. S. Watson, K. M. Watson, H. S. Pearson, M. F. Karn 1923. On the nest and eggs of the Common Tern (S. Fluviatilis). A third cooperative study. Biometrika 15: 294-345. Table LL on page 342.

(15) Using the same data (exhibit 32) calculate array trimeans in both directions. Plot the results, comparing with the two lines fitted to the thinned batch of 30 or 31.

(16) Age and auricular height of head for 2272 girls (see exhibit 33). Thin down to a batch of 22 or 23 (by spacing 100). Plot the resulting thinned batch.

(17) Fit a line for auricular height against age. (x vs. y) Plot lines and thinned points together.

(18) For the same data (exhibit 33) find and plot array trimeans of auricular height for each age. Plot array trimeans and the thinned=batch lines on the same sheet. How do they compare?

(19) Sir Francis Galton's measurements of reaction time to sight and head length (second laboratory series) for 4690 males (see exhibit 34). Replace by 23 or 24 points in 5 ways (by choosing 5 different starting points and spacing 200. Plot each batch separately. How well do you think the five sets of 23 or 24 agree? Also plot all 5 batches on the same sheet.

(20) Using the same data (exhibit 34) calculate array trimeans in both directions. Plot the results.

(21) Find, from other sources, a batch of at least 250 (x,y) data-sets that interests you and deserves treatment by the methods of this chapter. Treat it accordingly.

exhibit 33 of chapter 13

Age and auricular height of head in girls

Auricular height in millimeters**	Age in years* lo	7+	8+	9+	10+	11+	12+	13+	14+	15+	16+	17+	hi	Total
lo	6	6	3	3	5	8	3	8	1	4	1	.	1	49
111+	4	1	5	12	3	6	5	3	9	2	4	.	1	56
113+	1	4	3	10	8	6	9	4	3	5	5	1	.	59
115+	8	7	8	14	14	11	16	10	7	6	8	2	3	115
117+	13	9	7	10	23	15	18	13	9	11	6	4	4	142
119+	8	13	22	24	25	37	44	23	11	19	6	6	6	244
121+	11	9	19	25	29	34	41	32	21	15	13	9	7	265
123+	6	7	17	23	34	38	33	21	22	18	25	9	4	261
125+	1	6	19	18	33	39	40	32	23	26	14	12	12	265
127+	7	9	10	8	21	27	27	32	20	18	16	13	11	219
129+	.	.	6	9	17	16	20	39	25	29	16	11	9	197
131+	1	3	5	5	7	13	17	17	15	18	12	6	12	131
133+	.	1	.	7	8	10	13	8	5	16	7	7	6	88
135+	.	1	1	3	4	4	9	11	13	9	11	8	3	77
137+	.	.	.	3	2	2	10	4	5	14	6	3	3	52
hi	.	.	.	2	2	5	4	6	9	4	12	3	5	52
Total	66	76	125	177	235	261	309	263	198	214	162	95	91	2272

* 11+, for example, is at least 11 but not yet 12. Others similar.
** 111+, for example, is 110.25 to 112.25. Others similar.

----------supplementary detail for "lo" and "hi" cases (age,auric.ht), "+" implied----------

Once each: (3,115), (4,103), (5,103), (5,107), (5,111), (5,113), (5,127), (6,109), (6,125), (6,131), (7,107), (8,105), (9,105), (9,107), (9,109), (9,141), (9,143), (10,105), (12,105), (12,141), (14,109), (15,107), (16,109), (17,145), (18,111), (18,141), (18,143), (19,109), (19,123), (19,127), (19,135), (19,137), (19,141), (19,147), (20,115), (20,119), (20,125), (20,129), (20,131), (21,117), (21,125), (21,129), (21,131), (21,135), (22,123), (22,127).
Twice each: (4,117), (4,119), (4,121), (5,117), (5,119), (6,105), (8,109), (10,139), (11,105), (11,109), (11,141), (12,109), (13,105), (13,107), (13,139), (13,141), (13,143), (14,139), (15,139), (15,141), (16,141), (16,145), (17,139), (18,115), (18,135), (18,137), (19,111), (20,123). Three times: (5,115), (5,121), (5,123), (6,111), (6,123), (11,139), (12,139), (14,143), (15,109), (18,117), (18,119), (21,121). Four times: (6,115), (6,119), (10,109), (11,107), (13,109),(14,141), (16,139), (16,143), (18,121), (18,123), (19,131).

(cont'd.)

exhibit 33 (cont'd.)

Five times: (7,109). Six times: (6,121), (6,127), (18,131), (18,133). Seven times: (18,129). Nine times: (6,117), (18,127). Ten times: (18,125). (246=246, check!)

Source: K. Pearson 1905 and 1948. Mathematical Contributions to the theory of evolution - XIV. On the general theory of skew correlation and nonlinear regression. (Drapers Company Research Memoirs, Biometric Series, II) reprinted at pages 477-528 of Karl Pearson's Early Statistical Papers (1948) Cambridge University Press. Table III to face page 508.

exhibit 34 of chapter 13

Head length in inches	Reaction time in 1/100 second											Total
	lo	10	12	14	16	18	20	22	24	26	hi	
lo	.	1	.	2	4	9	8	2	1	.	1	28
7.1	.	2	1	4	2	11	13	7	.	3	.	43
7.2	.	1	1	6	14	22	37	11	7	3	2	104
7.3	1	5	1	10	37	43	58	25	10	5	1	196
7.4	.	5	5	23	67	76	122	41	24	13	2	378
7.5	2	10	9	29	115	143	148	53	22	13	6	550
7.6	1	10	10	30	131	191	176	76	58	13	6	702
7.7	.	12	15	49	138	201	196	56	45	12	7	731
7.8	.	12	15	39	145	177	215	78	42	14	8	745
7.9	.	7	5	27	121	146	138	49	24	5	3	525
8.0	.	6	8	19	71	83	81	28	19	5	4	324
8.1	.	4	6	9	39	52	63	25	11	6	3	218
8.2	.	1	4	7	17	22	31	8	5	2	.	97
8.3	.	.	2	1	6	11	9	5	1	.	.	35
hi	.	.	1	.	3	3	5	1	1	.	.	14
Total	4	76	83	255	910	1190	1300	465	270	94	43	4690

* All can be lowered 0.005. 14, for example, covers 13.00 to 14.99 and can be taken at 13.995.

** All can be raised 0.005. 7.9, for example, covers 7.66 to 7.95, and can be taken as 7.905.

----------supplementary detail: (head length,reaction time) for "lo" and "hi" cases----------

Once each: (7.5,6), (7.3,8), (7.5,8), (7.6,8), (7.0,10), (8.4,12), (8.4,14), (8.4,15), (6.9,20), (6.9,22), (7.0,22), (8.5,22), (7.0,24), (8.5,24), (7.0,28), (7.2,28), (7.4,28), (7.2,30), (7.6,30), (7.9,30), (8.1,30), (7.5,32), (7.8,32), (7.4,34), (7.8,36), (7.3,40), (7.7,44). Twice each: (7.0,14), (8.5,16), (8.5,18), (6.8,20), (8.6,20), (7.7,28), (7.9,28), (8.0,28), (8.1,28), (7.8,30), (8.0,30). Three times: (8.4,20). Four times: (7.0,16), (7.0,18), (7.8,28), (7.7,30). Five times: (6.9,18), (7.0,20), (7.5,28), (7.6,28). (Check: 89=89)

Source: G. E. Harmon 1926. On the degree of relationship between head measurements and reaction time to sight and sound. Biometrika 18: 207-220. Table IV on page 215. (Used by L. H. C. Tippett as table 8.1d on page 223 of the 4th edition (1952) of The Methods of Statistics, John Wiley, New York.)